Technological University Dublin
ARROW@TU Dublin

Conference papers | School of Computer Science

2023

# Dynamic Influence Diagram-Based Deep Reinforcement Learning Framework and Application for Decision Support for Operators in Control Rooms

Joseph Mietkiewicz
*Technological University Dublin, Ireland*, joseph.mietkiewicz@tudublin.ie

Ammar N. Abbas
*Technological University Dublin, Ireland*, ammar.abbas@tudublin.ie

Chidera Winifred Amazu
*Politecnico di Torino, Italy*, D21127591@mytudublin.ie

*See next page for additional authors*

Follow this and additional works at: https://arrow.tudublin.ie/scschcomcon

Part of the Computer Engineering Commons

## Recommended Citation

Mietkiewicz, Joseph; Abbas, Ammar N.; Amazu, Chidera Winifred; Madsen, Anders L.; and Baldissone, Gabriele, "Dynamic Influence Diagram-Based Deep Reinforcement Learning Framework and Application for Decision Support for Operators in Control Rooms" (2023). *Conference papers*. 410.
https://arrow.tudublin.ie/scschcomcon/410

This Conference Paper is brought to you for free and open access by the School of Computer Science at ARROW@TU Dublin. It has been accepted for inclusion in Conference papers by an authorized administrator of ARROW@TU Dublin. For more information, please contact arrow.admin@tudublin.ie.

This work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.
Funder: European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement no. 955901

## Authors

Joseph Mietkiewicz, Ammar N. Abbas, Chidera Winifred Amazu, Anders L. Madsen, and Gabriele Baldissone

This conference paper is available at ARROW@TU Dublin: https://arrow.tudublin.ie/scschcomcon/410

# Dynamic Influence Diagram-Based Deep Reinforcement Learning Framework and Application for Decision Support for Operators in Control Rooms

Joseph Mietkiewicz

*Hugin Expert A/S /TU Dublin, Denmark.*
*Food Science Environmental Health, Technological University Dublin, Ireland.*
*E-mail: D21127042@mytudublin.ie*

Ammar N. Abbas

*Department of Computer Science, Technological University Dublin, Ireland.*
*Data Science, Software Competence Center Hagenberg, Austria. E-mail: ammar.abbas@scch.at*

Chidera Winifred Amazu

*Safety, Reliability and Risk Centre (SAfeR), Department of Applied Sciences and Technology,*
*Politecnico di Torino, Italy. E-mail: chidera.amazu@polito.it*

Anders L. Madsen

*Hugin Expert A/S, Department of computer science Aalborg University, Denmark. E-mail: anders@hugin.com*

Gabriele Baldissone

*SAfeR - Department of Applied Sciences and Technology, Politecnico di Torino, Italy.*
*E-mail: gabriele.baldissone@polito.it*

In today's complex industrial environment, operators are often faced with challenging situations that require quick and accurate decision-making. The human-machine interface (HMI) can display too much information, leading to information overload and potentially compromising the operator's ability to respond effectively. To address this challenge, decision support models are needed to assist operators in identifying and responding to potential safety incidents. In this paper, we present an experiment to evaluate the effectiveness of a recommendation system in addressing the challenge of information overload. The case study focuses on a formaldehyde production simulator and examines the performance of an improved Human-Machine Interface (HMI) with the use of an AI-based recommendation system utilizing a dynamic influence diagram in conjunction with reinforcement learning. The preliminary results indicate the potential of these methods to aid operators in decision-making during challenging situations and enhance process safety in the industry.

*Keywords*: Decision support, Dynamic influence diagram, Reinforcement learning, Process safety, Intervention procedures, Human-machine interface

## 1. Introduction

In a chemical process control room, operators are responsible for monitoring the process and responding to deviations. To do so they have to handle the available information and follow a set of procedures in order to restore the process to normal. However, human errors can occur due to overload or inadequate training. To address these issues, optimal procedures are required to assist operators in their tasks. We develop a system that provides operators with the appropriate and optimal procedure in real-time using a dynamic influence diagram combined with reinforcement learning. This system detects anomalies and provides operators with the reason for the deviation, as well as the specific procedure to follow in each case updated with the current state of the system. This framework is then tested on a simulator of formaldehyde production. Three deviation scenarios are designed and presented to two groups of participants. The first group has to respond to the scenario without the recommendation system and

Proceedings of the 33rd European Safety and Reliability Conference.
*Edited by* Maria Chiara Leva, Hector Diego Estrada Lugo, Edoardo Patelli and Luca Podofillini
Copyright © 2023 by ESREL2023 Organizers. *Published by* Research Publishing, Singapore
ISBN: 981-973-0000-00-0 :: doi: 10.3850/981-973-0000-00-0_output

the second group with. The efficacy of the recommendation system is then evaluated based on its ability to assist the participants in handling deviant scenarios. The objective of this experiment is to determine if this approach reduces the operator's workload and enhances situational awareness.

## 2. Background

The experiment is made to optimize the display of the procedure to the operator. For this purpose, we use artificial intelligence to display directly to the operator the right procedure update with the live value of the process. For this purpose, a dynamic influence diagram combined with reinforcement learning is used.

### 2.1. *Procedures*

Procedures are supporting tools that guide control operators to manage process plant operations and safety. These process operations can be; the start-up and shut-down of a plant, normal operations, emergency operations, and more. In a study done by Bakar et al. (2017), to identify process safety management elements that are main accident contributors (from amongst $770\,\mathrm{major}$ accident cases) in chemical process plants, Operating procedures were found to be the second most contributors to accidents in this plants at $17\,\%$ just after process failures at $19\,\%$. Similarly, findings in the nuclear industry have also shown that operators' non-compliance behaviour due to procedures is one of the leading causes of accidents in this industry Park and Jung (2003). As one of the process safety management elements, these procedures are critical to ensuring safety and should be paid close attention to beyond the initial stages of design and installation.

Several constraints the procedures pose leading to these non-compliance behaviours have been highlighted in industrial studies. These constraints, related to inaccuracy or incompleteness and complexity of the procedures, have led to non-compliance issues by operators, which is one of the reasons for the accidents Park and Jung (2003). Also, there have been studies on constraints related to design. That is, the design of these procedures (medium of presentation and the presentation style) and how they impact the operators' workload, especially in emergencies, seeing that some industries have moved over years from paper-based, screen-based and, in some cases, computerised procedures Gao et al. (2013),Xu et al. (2008). The ease of use of the procedures, their ability to support operators' situational awareness and reduce workload within the available time in abnormal situations are goals to be attained by decision-makers if safety is to be ensured. In this work, a screen-based procedure (following best practices in procedural formatting - HPOG) added or not with AI-based support is developed to compare and assess the impact of different procedural support mediums on process control room operators during an intervention in process safety scenarios.

### 2.2. *AI-based Recommendation System*

The AI-based recommendation system is using both Deep Reinforcement Learning (DRL) and a dynamic influence diagram. The DRL is trained on the simulator in an online setting by interacting with the environment and observing process behavior through its actions. The dynamic influence diagram is constructed from expert knowledge consisting of the physical equation of the plants. Those models are combined to detect deviance in the process and recommend the best action to the operator by providing a detailed procedure with specific values adapted to the current situation.

#### 2.2.1. *Dynamic Influence diagram*

In the realm of decision-making under uncertainty, influence diagrams Howard (1983) have emerged as an effective modeling framework. These diagrams provide a natural way to capture the semantics of decision-making with minimal clutter and confusion for the decision-maker Shachter and Peot (1992). It is perfect to model uncertainty in the process and provide the best recommendation. An influence diagram is a graphical probabilistic model that provides a framework for optimal decision-making under uncertainty. Dynamic Influence diagrams (DID) can be used to monitor the process over time by modeling the dynamics of the system and the influence of various

factors on the system's behavior. DID can also be used to predict the future behavior of the system by simulating different scenarios and observing their impact on the overall system. This allows for better decision-making based on the potential outcomes of different choices. Kjærulff et al. (2013).

In this study, the physical model of the plant is used to build the dynamic influence diagram. The different nodes can represent:

(1) values, such as temperature, pressure, etc.
(2) temporal clones of values, such as temperature, and pressure, representing the influence of the past.
(3) failure, such as control valve failure.
(4) decisions, such as set points.
(5) costs associated with a particular event.

Each of these nodes is linked to other nodes through causal links. There is an associated cost called utility with each possible action of the decision node. The model outputs to the operator the action with the maximum expected utility.

For example, in the case of choosing the right flow of nitrogen to avoid overpressure in a tank. If $A$ is a set point for a flow with options $0-1,\ldots,6-7$ $(Nm^3/h)$, $H$ is a specific alarm link to the underpressure of the tank with states $False, True$, and $\epsilon$ is a set of observations in the form of evidence, then we can compute the utility of each outcome of the hypothesis and the expected utility of each action. The utility of an outcome $(0-1, True)$ is $U(0-1, True)$, where $U(\cdot)$ is our utility function. The expected utility of performing action setting $4-5$ $Nm^3/h$ as a nitrogen flow is calculated by

$$EU(4-5) = U(4-5, True)P(True|\epsilon) + U(4-5, False)P(False|\epsilon)$$

#### 2.2.2. *Using Dynamic Influence Diagram*

The model is dynamic to monitor the process over time. The model assesses the state of the system at the current time and predicts the future state for numerous subsequent time steps. The dynamic influence diagram serves three distinct purposes:

(1) Detecting anomalies
(2) Predicting the future state of the system
(3) Determine an optimal decision.

The model is built using expert knowledge in the form of the physical equation of the process. The detection of anomalies uses the present data and can detect for example a deviation from a set-point. The model is able to predict the future state of the system for the next time steps. According to this prediction, an ideal scenario can be built and presented to the operator. This recommendation is based on the procedure associated with the detected anomaly. The procedure is adapted to provide concise instruction and precise values to set according to the system's current state to the operator.

#### 2.2.3. *Specialized Reinforcement learning Agent (SRLA)*

**Reinforcement Learning** (RL) is a type of machine learning in which an agent learns to make decisions by interacting with an environment and receiving feedback in the form of rewards or penalties. The goal of the agent is to learn a policy, or a mapping of states to actions, that maximizes its cumulative reward over time Sutton and Barto (2018).

The equation of reinforcement learning can be represented using the Q-learning algorithm, which is a model-free reinforcement learning algorithm. In Q-learning, an agent learns to make decisions by iteratively updating a Q-value function that estimates the expected cumulative reward for taking a particular action in a particular state. The Q-value function can be represented as follows:

$$Q(s,a) = E[r + \gamma * \max_{a'} Q(s', a')] \quad (1)$$

DRL extends Q-learning by using deep neural networks to estimate the Q-value function. The neural network takes the state as input and produces the Q-value estimates for all possible actions as output. The Q-value function is then updated by minimizing the difference between the predicted Q-values and the actual Q-values obtained from the rewards received by the agent during training. The equation for updating the Q-value function in deep reinforcement learning can be represented as follows:

$$Q(s,a) = Q(s,a) + \alpha(r + \gamma * \max_{a'} Q(s',a') - Q(s,a)) \quad (2)$$

In both equations, the symbol $\gamma$ represents the discount factor, $r$ represents the immediate reward received by the agent, $s$ represents the current state of the agent, $a$ represents the action taken by the agent, $s'$ represents the next state of the agent, and $a'$ represents the action taken in the next state. The symbol $\alpha$ represents the learning rate, which determines the step size of the Q-value updates.

**Specialized Reinforcement Learning Agent** (SRLA) combines the advantages of probabilistic modeling (such as dynamic influence diagram) and Deep Reinforcement Learning (DRL) Abbas et al. (2022) as shown in fig. 1. It allows the DRL agent to specialize in a specific case within the environment where the DRL agent is most required (such as process abnormalities). Hence, increasing training efficiency and reducing data inefficiency.

In the figure, $P(s_t)$ represents the probability of being in a particular state, $x^*(s_t)$ represents the specialized state on which the DRL agent is activated and the system state $S$ is filtered to provide the information regarding that particular state. The optimal control strategy $\pi$ is then recommended to the operator.

SRLA modifies eq. (2) as shown in eq. (3):

$$Q(x^*,a) = Q(x^*,a) + \alpha(r + \gamma * \max_{a'} Q(x^{*'},a') - Q(x^*,a)) \quad (3)$$

where $x^*$ represents the specialized state on which the DRL agent is activated and trained and is identified through a probabilistic model.

## 3. Framework

In our proposed AI recommendation system framework with the Human-in-the-Loop (HITL) setting, we use Multi Specialized Reinforcement Learning Agent (M-SRLA) setting. The multi-agents in this setting act independently and only a specific agent is activated to suggest the optimal control strategy to the operator in case of the identified process abnormality through the influence diagram as shown in fig. 4. We name this framework Human-Centered Artificial Intelligence for Safety-Critical systems "HAISC".

### 3.1. *Procedure and Dynamic Influence Diagram*

The procedure in this study is written as a hierarchical rule-based task representation format and is further presented on screen for the operators. Each procedure is written under three broad task contexts; troubleshooting, control, and evaluation. An example is shown in fig. 2. The recommendation of the influence diagram is based on the procedure. The influence diagram detects the anomaly and recommends the procedure associated with it. The procedure is simplified by directly providing the action to do without troubleshooting and control. In case of different anomalies at the same time, the influence diagram chose the best action to do according to their utility. The specific value to set in the procedure is then specified jointly with reinforcement learning. In this way, the procedure is adapted to the specific situation of the system and provide the operator with a complete and concise set of instruction.

### 3.2. *Dynamic Influence Diagram and Reinforcement learning*

The approach uses an influence diagram to identify the specific failure in the system. Once the procedure associated with the failure is identified, the necessary step of the procedure according to the current situation is displayed to the operator. If the procedure requires continuous values to be set, reinforcement learning is employed to specify the appropriate value based on the current state of the system. The influence diagram provides a value in the form of an interval due to the discretization of the data. If the value obtained from the reinforcement learning falls outside of this interval, only the procedure along with the interval is displayed to the operator. This is done to ensure the interpretability and safety of the overall system. All in all the influence diagram models the system globally and the reinforcement learning answers

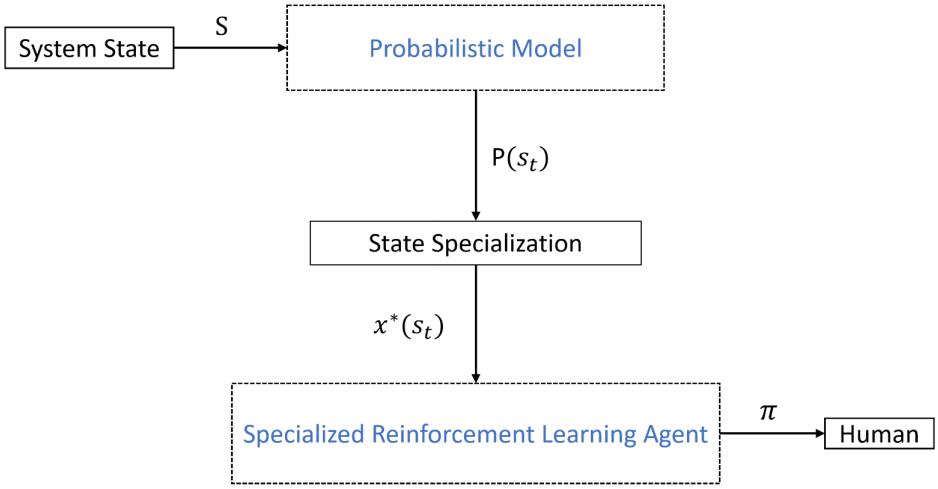


Fig. 1. Specialized Reinforcement Learning Agent (SRLA) Abbas et al. (2022)

precisely local issues.

**Algorithm 1** Influence Diagram-based Procedure Recommendation Algorithm

**Input:** System parameters
**Output:** Procedure with necessary steps and specific values
Detect anomalies
**if** *multiple anomalies* **then**
  Choose a procedure with maximum expected utility;
**else**
  Use procedure link to the anomaly;
**end**
**if** *specific value needs to be set* **then**
  Use DID to find the best interval I
  Call DRL to determine the value a
  **if** $a \in I$ **then**
    *Use value;*
  **else**
    *Use interval;*
  **end**
**end**

### 3.3. *Recommendation*

The operator is provided with recommendations that include the identified faults and the appropriate procedures for restoring the system to its normal state. Traditionally, these procedures were developed by experts. However, with the advent of the AI framework, the procedures have been simplified and specific values are now provided based on the current state of the system. This eliminates any extraneous or irrelevant steps, allowing the operator to quickly and easily follow the complete set of recommendations. fig. 3 depicts an example of a simplified procedure compared to the traditional one.

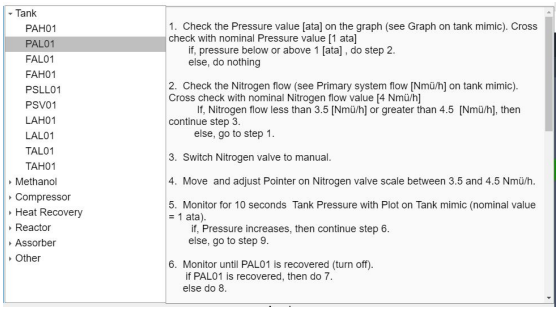
Tank
PAH01
PAL01
FAL01
FAH01
PSLL01
PSV01
LAH01
LAL01
TAL01
TAH01
Methanol
Compressor
Heat Recovery
Reactor
Assorber
Other
1. Check the Pressure value [ata] on the graph (see Graph on tank mimic). Cross check with nominal Pressure value [1 ata]
if, pressure below or above 1 [ata] , do step 2.
else, do nothing
2. Check the Nitrogen flow (see Primary system flow [Nmü/h] on tank mimic). Cross check with nominal Nitrogen flow value [4 Nmü/h]
If, Nitrogen flow less than 3.5 [Nmü/h] or greater than 4.5 [Nmü/h], then continue step 3.
else, go to step 1.
3. Switch Nitrogen valve to manual.
4. Move and adjust Pointer on Nitrogen valve scale between 3.5 and 4.5 Nmü/h.
5. Monitor for 10 seconds Tank Pressure with Plot on Tank mimic (nominal value = 1 ata).
if, Pressure increases, then continue step 6.
else, go to step 9.
6. Monitor until PAL01 is recovered (turn off).
if PAL01 is recovered, then do 7.
else do 8.

Fig. 2. First steps of the procedure for one alarm

1. Open Graph on tank mimic
2. Switch the Nitrogen valve to a manual.
3. Move and adjust the Pointer on the Nitrogen valve scale at 3.9 $Nm^3/h$.

Fig. 3. Simplified procedure recommended to the operator

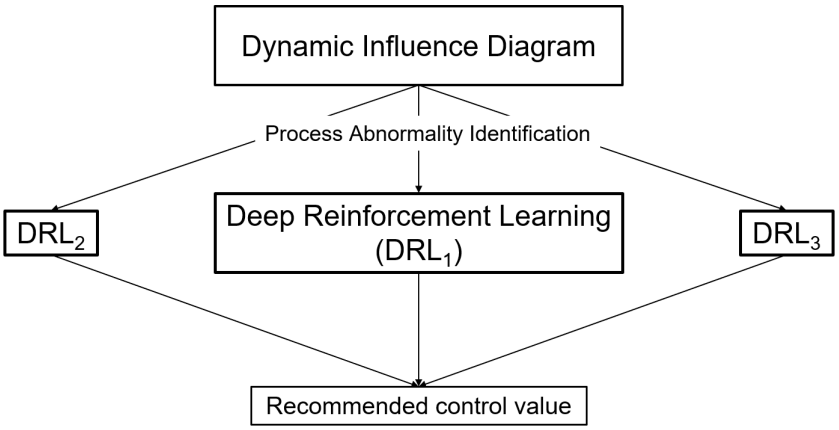


Fig. 4. Framework used for the case studies

## 4. Experimental study

### 4.1. *Simulator*

The simulator is one of the productions of formaldehyde operating the partial oxidation of methanol with air. The details of the simulator can be found in Amazu and all (2023). The operator is responsible for monitoring the process and correcting any deviations. fig. 5 shows the main screen of the simulator, and for this preliminary study, three scenarios have been created. In the first scenario, a control valve failure in the tank section occurred. The operator must open the tank part of the system, switch to manual mode, and adjust the set point of the flow value. In the second scenario, a valve failure in the tank section occurred. In this scenario, the operator must switch to the backup system and adjust the pump power due to a delay in the nitrogen flow coming from the backup system. The third scenario simulated a failure of temperature indicator control that has an impact on the reactor. The operator needs to maintain the right temperature in the reactor. The first group of participants has all the procedures available on the screen, while the second group has additional support from the recommendation system which provides a concise live procedure.

### 4.2. *Physiological measurement*

Physiological measurement is used to have an objective measurement of the operator's workload and situation awareness and have a better insight into the recommendation system's benefit.

#### 4.2.1. *Heart rate*

By monitoring the operator's heart rate (HR) and heart rate variability (HRV), before and after using the recommendation system, we can determine whether the system is reducing or increasing their stress and workload levels Gao et al. (2013), Muhajir et al. (2021). If the recommendation system is effective, we would expect to see a decrease in the operator's workload and stress levels after using the system compared to before. However, it is important to note that a decrease in heart rate doesn't necessarily mean that the recommendation system is effective. There are other factors that could contribute to a decrease in heart rate, such as taking a break, delegating tasks, or even just the passage of time. Therefore, it is important to combine these physiological measures with other metrics such as task completion time and error rate to determine the effectiveness of the recommendation system. Monitoring the operator's HR, HRV, and EDA, can provide valuable insights into their workload and the effectiveness of the recommendation system.

#### 4.2.2. *Eye tracker*

Eye tracking is a valuable technique for monitoring participants during experiments. By tracking eye movements, it is possible to gain more insights into the participant's cognitive states of attention, mental workload, and situational awareness Hinss

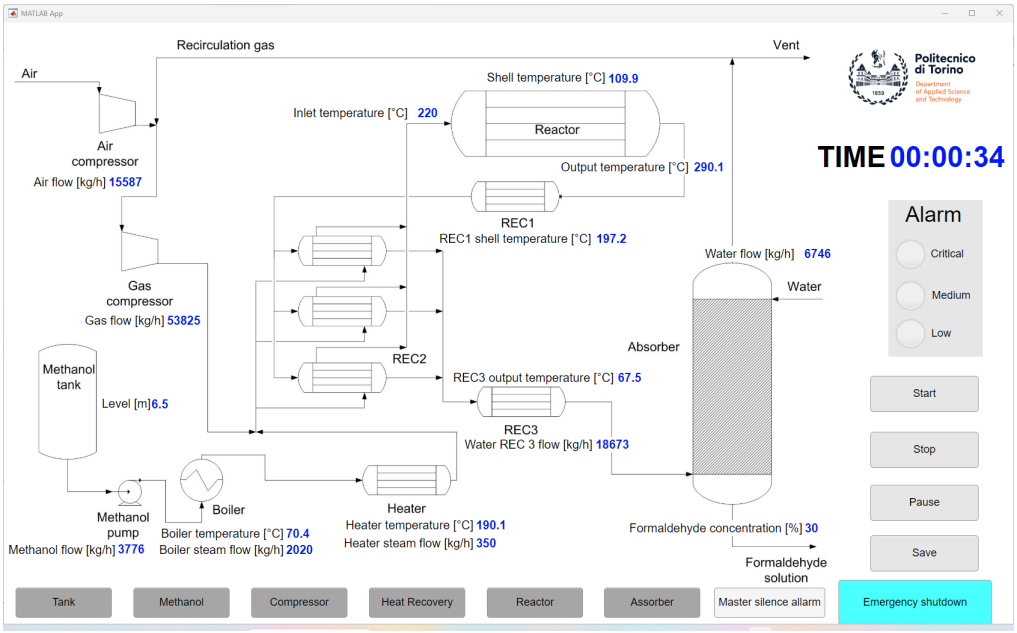


Fig. 5. Simulator main screen

et al. (2022), Amazu et al. (2022). During the experiment, the eye tracker monitors the participant's eye movement as they interact with the recommendation system on the human-machine interface and respond to the different scenarios. This information is analyzed to determine which support types (screen or AI) are most effective in guiding the participant's decision-making process. Additionally, eye tracking can reveal when and where the participant experiences difficulty or confusion, allowing researchers to identify areas that require improvement.

## 4.3. *Questionnaire*

### 4.3.1. *Workload*

NASA-TLX (National Aeronautics and Space Administration-Task Load Index) is a widely used subjective rating scale that measures the perceived workload and mental effort required to complete a task. During the experiment, NASA-TLX is used to assess the participant's subjective workload after operating in a scenario. Participants rate their overall perceived workload in the scenario based on six different dimensions, including mental demand, physical demand, temporal demand, performance, effort, and frustration. The result compares the workload of people with and without recommendations during the scenarios.

### 4.3.2. *Situation awareness*

The Situation Awareness Rating Technique (SART) is a subjective rating technique used to assess situational awareness post-trial. It measures the situational awareness of test participants in ten dimensions; attention, information quality, and more. During the experiment, SART is completed post-scenario by the participants to assess their overall situational awareness. The use of SART in this experimental study allows us to study the variation in situational awareness between the two groups.

### 4.3.3. *Recommendation system*

Questionnaires are given to the participant to assess the efficiency of the recommendation system. The questionnaire is designed to gather feedback from the participants on the ease of use, clarity, and relevance of the recommendations provided by the system. The questionnaire includes questions related to the participant's perception of the recommendation system's ability to detect faults, the accuracy of the recommendations provided,

and the system's usefulness in guiding decision-making during the deviance scenarios.

## 5. Preliminary result

A pilot study was conducted to evaluate the performance of the recommendation system. It was with 2 participants without the recommendation system and 2 with. It did not allow us to obtain sufficiently significant statistical data. But we observed that the group with the recommendation system acts faster but tends to follow the recommendation blindly. In this case, the recommendation system tends to reduce the workload but also to reduce situational awareness. These promising results suggest that the recommendation system effectively reduces workload but decreases situational awareness during process safety decision-making. Further research with larger participant samples is needed to confirm these findings and to optimize the design of the recommendation system for wider application.

## 6. Conclusion

The presentation of the procedure to the operator is crucial. Providing the operator with the right and simplified procedure that adapts to the current system state has shown promising results in assisting the operator's work. Thanks to the collaboration between the influence diagram and reinforcement learning, the recommended procedure is interpretable and based on expert knowledge. This system would be particularly advantageous for operators in situations of overload, stress, or inexperienced operators.

### Acknowledgement

This work has been done within the Collaborative Intelligence for Safety-Critical Systems project (CISC). The CISC project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement no. 955901.

## References

Abbas, A. N., G. C. Chasparis, and J. D. Kelleher (2022). Interpretable input-output hidden markov model-based deep reinforcement learning for the predictive maintenance of turbofan engines. In *Big Data Analytics and Knowledge Discovery: 24th International Conference, DaWaK 2022, Vienna, Austria, August 22–24, 2022, Proceedings*, pp. 133–148. Springer.

Amazu, B. and all (2023). Analysing "human-in-the-loop" for advances in process safety: a design of experiment in a simulated process control room.

Amazu, C. W., M. Demichela, and D. Fissore (2022). Human-in-the-loop configurations in process and energy industries: a systematic review. In M. C. Leva, Patelli Edoardo, Podofillini Luca, and Wilson Simon (Eds.), *Proceedings of the 32nd European Safety and Reliability Conference (ESREL 2022)*, pp. 3234–3241. Research Publishing, Singapore.

Bakar, H. T. A., P. H. Siong, C. K. Yan, K. Kidam, M. W. Ali, M. H. Hassim, and H. Kamarden (2017). Analysis of main accident contributor according to process safety management elements failure. *Chemical Engineering Transactions 56*, 991–996.

Gao, Q., Y. Wang, F. Song, Z. Li, and X. Dong (2013). Mental workload measurement for emergency operating procedures in digital nuclear power plants. *Ergonomics 56*(7), 1070–1085.

Hinss, M. F., A. M. Brock, and R. N. Roy (2022). Cognitive effects of prolonged continuous human-machine interaction: The case for mental state-based adaptive interfaces. *Frontiers in Neuroergonomics 3*.

Howard, R. A. (1983). *Readings on the principles and applications of decision analysis: General collection*, Volume 1. Strategic Decisions Group.

Kjærulff, U., A. Madsen, and B. Networks (2013). Influence diagrams: a guide to construction and analysis.

Muhajir, D., F. Mahananto, and N. A. Sani (2021). Stress level measurements using heart rate variability analysis on android based application. *Procedia Computer Science 197*(2021), 189–197.

Park, J. and W. Jung (2003). The operators' non-compliance behavior to conduct emergency operating procedures - Comparing with the work experience and the complexity of procedural steps. *Reliability Engineering and System Safety 82*(2), 115–131.

Shachter, R. D. and M. A. Peot (1992). Decision making using probabilistic inference methods. In *Uncertainty in Artificial Intelligence*, pp. 276–283. Elsevier.

Sutton, R. S. and A. G. Barto (2018). *Reinforcement learning: An introduction*. MIT press.

Xu, S., F. Song, Z. Li, Q. Zhao, W. Luo, X. He, and G. Salvendy (2008). An ergonomics study of computerized emergency operating procedures: Presentation style, task complexity, and training level. *Reliability Engineering and System Safety 93*(10), 1500–1511.